Whole Genome Analysis of Functional Protein Binding Sites and DNA Methylation: Application to p53 and Low Dose Ionizing Radiation.

Krassimira Botcheva, John J. Dunn and Carl W. Anderson
Biology Department, Brookhaven National Laboratory, Upton, NY 11973, USA

The effects of exposure to low doses of ionizing radiation on humans results largely from changes in gene expression mediated by the activation of sequence-specific DNA binding proteins (transcription factors) as well as changes to other chromosomal proteins and perhaps to DNA. To develop a molecular understanding of the consequences of exposures to low doses of ionizing radiation, it will be necessary to understanding where radiation-activated transcription factors bind in whole genomes and how radiation induces changes in factor binding and chromosome structure. We have developed a method, PE-SACO (Paired-end Serial Analysis of Chromatin Occupancy)[1], for profiling the functional chromatin binding sites of proteins in cells across whole genomes. PE-SACO represents an improved version of our original SACO technique, which was used to identify the binding sites of the CREB transcription factor in the whole genome of rat cells.[2] SACO combines chromatin immunoprecipitation with Genome Signature Tags[3], a technique that associates 20/21 bp DNA sequences (TAGs) with molecular events to identify and quantify functional protein binding sites in the chromatin of whole genomes. PE-SACO captures the sequence of diTAGs, 20/21 bp sequences from each of both ends of small, affinity-selected DNA fragments through the use of a special vector (pBEST) that has opposing *Mme* I sites positioned on both sides of the cloning site. After cleaving a cloned fragment library with *Mme* I, the vector with the 20/21 bp tags is recircularized by ligation with a linker fragment containing a flanking bar-code sequence. The diTAGs representing the paired-ends of selected DNA fragments are then generated by PCR amplification or cleavage with appropriate restriction enzymes. The size of the diTAGs, approximately 60 bp, is ideal for sequencing by new technologies such the Life Sciences 454 Genome Sequencer, which can deliver up to 800,000 diTAG sequences in 5.5 hr[4]; incorporation of sequence bar-codes allows several libraries to be combined for sequencing, further increasing the efficiency and decreasing costs. A somewhat similar technique, PET (Paired-end Tags), was independently developed by Wei et al.[5]

PE-SACO can be used to identify the binding sites for transcription factors in whole genomes, or to query chromatin structure by, e.g. identifying epigenetic marks such as posttranslational modifications to histones or DNA methylation sites. In collaboration with Gerald Pfeifer's group, we are developing PE-SACO for the analysis of changes in DNA methylation status. Several methylation-specific DNA binding proteins have been cloned and expressed and are being evaluated for their ability to query genomes for changes in the state of their CpG methylation.

The p53 tumor suppressor is a transcription factor that is activated in response cellular stresses including DNA double-strand breaks resulting from ionizing radiation. p53 binds to the consensus response element, 5'-RRRCWWGYYY (N = 0-14) RRRCWWGYYY-3', and regulates the expression, both positively or negatively, of ~1500 genes, but until recently, only a few of its response elements had been identified and characterized. Furthermore, p53-mediated activation or repression of transcription depends on interactions with other factors including histone acetyltransferases (HATs) and deacetylases (HDACs). Radiation-induced epigenetic changes (e.g. changes in histone modifications and/or DNA methylation) also affect cellular responses. Activation of p53 by low doses of ionizing radiation results in a transient cell cycle arrest and the induction of DNA repair enzymes but may also lead to premature senescence or the induction of apoptosis.

As an initial step towards characterizing the role of p53, we have created PE-SACO diTAG libraries from human IMR-90 fibroblasts exposed to no, 0.1 Gy or 8 Gy of ionizing radiation. In constructing these libraries, we implemented a bar-coding system utilizing fully self-complementary *Bcl* I linkers with the general structure: 5'-pXYZTGATCAZ'Y'X'NN-3' (where X, Y, and Z are any nucleotide and X', Y', and Z' are their complements). As for our original *Bcl* I linker, these bar-coding linkers are positioned between each diTAG sequence, and they have a fixed length, which facilitates parsing of individual flanking TAGs. The length of diTAGs with bar-coding (60 bp vs 54 bp) is still well within the limits of accurate sequencing on the Life Sciences 454 instrument. The GATC core of the *Bcl* I

recognition site is present in every DNA molecule, which provides a potential aid for sequence (peak) calling. Bar-coding the different libraries represents a significant improvement in PE-SACO since it allows multiplexing of libraries not only during sequencing but also during processing for all steps following ligation with the bar-code linkers.

Our p53 libraries are in the process of being sequenced by DOE's Joint Genome Institute In future studies we plan to extend our chromatin characterization to include changes in DNA methylation before and after exposure to ionizing radiation and/or oxidative DNA damage. We will describe additional improvements to the PE-SACO technology and show how it may be used in a systems approach to address other issues relevant to DOE's missions in energy security and bioremediation.

References

1. Dunn, J., McCorkle, S.R., Everett, L., and Anderson, C.W. 2006. Paired-end genomic signature tags: a method for the functional analysis of genomes and eipgenomes. in *Genetic Engineering: Principles and Methods* (ed. J.K. Setlow), p. (in press). Springer Science+Business Media, Inc., New York.
2. Impey, S., McCorkle, S.R., Cha-Molstad, H., Dwyer, J.M., Yochum, G.S., Boss, J.M., McWeeney, S., Dunn, J.J., Mandel, G., and Goodman, R.H. 2004. Defining the CREB regulon; a genome-wide analysis of transcription factor regulatory regions. *Cell* **119**(7): 1041-1054.
3. Dunn, J.J., McCorkle, S.R., Praissman, L.A., Hind, G., Van Der Lelie, D., Bahou, W.F., Gnatenko, D.V., and Krause, M.K. 2002. Genomic signature tags (GSTs): a system for profiling genomic DNA. *Genome Res* **12**(11): 1756-1765.
4. Margulies, M., Egholm, M., Altman, W.E., Attiya, S., Bader, J.S., et al. 2005. Genome sequencing in microfabricated high-density picolitre reactors. *Nature* **437**(7057): 376-380.
5. Wei, C.-L., Wu, Q., Vega, V.B., Chiu, K.P., Ng, P., Zhang, T., Shahab, A., Yong, H.C., Fu, Y., Weng, Z., Liu, J., Zhao, X.D., Chew, J.-L., Lee, Y.L., Kuznetsov, V.A., Sung, W.-K., Miller, L.D., Lim, B., Liu, E.T., Yu, Q., Ng, H.-H., and Ruan, Y. 2006. A global map of p53 transcription-factor binding sites in the human genome. *Cell* **124**(1): 207-219.